Random Peptide Libraries: A Source of Specific Protein Binding Molecules

JAMES J. DEVLIN,* LUCY C. PANGANIBAN, PATRICIA E. DEVLIN

Libraries of random peptide sequences were constructed and screened to identify peptides that specifically bind to proteins. In one of these about 2×10^7 different 15-residue peptide sequences were expressed on the surface of the coliphage M13. Each phage encoded a single random sequence and expressesd it as a fusion complex with pIII, a minor coat protein present at five molecules per phage. Phage encoding nine different streptavidin-binding peptide sequences were isolated from this library. The core consensus sequence was His-Pro-Gln and binding of these phage to streptavidin was inhibited by biotin. This type of library makes it possible to identify peptides that bind to proteins (or other macromolecules) that have no previously known affinity for peptides.

SMALL MOLECULES THAT INTERACT with proteins (enzyme inhibitors or receptor agonists and antagonists, for example) are used in research and in clinical therapy. These small molecules can be developed by rational design or isolated by screening large numbers of naturally occurring or synthetic compounds. Frequently there is not enough information available for design, and assembling and screening a large library of compounds has been time consuming and expensive. We now describe the use of biological expression systems to facilitate both the production of large libraries of random peptide sequences and the screening of these libraries for peptide sequences that bind to particular proteins.

A major goal has been to increase the number of sequences that can be screened at one time. We produced the libraries by cloning synthetic DNA that encoded random peptide sequences into *Escherichia coli* expression vectors. For our first library, we used the phage λgt11 (*1*) to express 4×10^6 random 15-residue peptide sequences as fusion proteins with β-galactosidase. Although we identified fusion proteins that were recognized by an antibody to a linear epitope of the feline leukemia virus, this system was limited to screening about 10^6 peptide sequences at a time. To increase the number of sequences that could be screened by potentially several orders of magnitude, we then investigated a filamentous phage expression system.

In the filamentous phage system, foreign peptide sequences can be expressed on the surface of infectious phage (*2*), thereby conferring two significant benefits. First, since these phage can be applied to affinity matrices at concentrations well over 10^{13} phage per milliliter, large numbers of phage can be screened at one time. Second, since each infectious phage encodes the random sequence expressed on its surface, if a particular phage is recovered from an affinity matrix in low yield, the phage can be amplified by another round of infection. The group of almost identical *E. coli* filamentous phages M13, fd, and f1 have five copies of a minor coat protein, pIII, on the surface (*3*). Parmley and Smith have shown that several foreign epitopes can be expressed at the NH_2-terminal end of pIII and furthermore, phage bearing one of these epitopes could be recovered from a large excess of phage lacking this epitope (*4*). They also suggested that if it were possible to express a large number of random peptide sequences, this system could be used to characterize the epitopes recognized by antibodies.

We have produced a large library of phage expressing random 15-residue peptide sequences as gene III fusion proteins; the accompanying article by Scott and Smith (*5*) describes a library of random six-residue sequences also expressed as pIII fusion proteins. We constructed the new expression vector, M13LP67, by making noncoding base changes to introduce restriction enzyme sites into gene III of M13mp19 and by placing a β-lactamase gene in the polylinker 3′ of the *lac* promoter (*6*). Ligation of complimentary oligonucleotides into the new sites in gene III led to mature pIII fusion proteins with the predicted NH_2-terminal sequence of Ala-Glu-Xxx_{15}-Pro_6-Ala-Glu (where Xxx represents random amino acid residues) (*7*). We retained the first two residues of the wild-type mature NH_2-terminus (Ala-Glu) because they may influence processing by signal peptidase, which is required to produce infectious phage (*8*). Both the frequency of termination codons and the variation in the number of codons for each amino acid residue was reduced by using $(NNS)_{15}$ to encode the 15 random residues (where N is a mixture of G, A, T, and C and S is a mixture of G and C). NNS encodes all 20 amino acids but will produce only one of the three termination codons. In the portion of the oligonucleotide complementary to $(NNS)_{15}$, we used deoxyinosine because of its relatively unselective base pairing ability (*9*). Polyproline tends to adopt an extended structure (*10*); thus, the random residues were followed by a spacer consisting of six proline residues, which was inserted to make the random peptides more accessible by moving them away from the rest of the protein. Since pIII function is required to infect *E. coli* (*11*), the Ala-Glu sequence was repeated after the proline linker to retain the entire uninterrupted sequence of pIII.

After these oligonucleotides were ligated into the M13LP67 expression vector, DNA was introduced into *E. coli* by electroporation, and these cells were plated with fresh *E. coli* cells (*12*). Successful introduction of the vector into a host cell resulted in a plaque of slow-growing phage-producing cells. Modification of pIII may reduce infectivity (*4*), and hence some phage replicating in a competitive environment may become underrepresented in the library. Such competition among the phage was avoided by plating transfected (by electroporation) cells at a density such that the plaques just touched one another (about 400 plaques per square centimeter).

After harvesting phage from 2.8×10^7 plaques, we determined what percentage of phage had the potential to express random peptide sequences. Phage that encoded random sequences contained an extra 69 base pairs (bp) in the 5′ end of gene III. Using polymerase chain reaction (PCR) to amplify the DNA from this region, we found that the 69-bp insert was present in 71 percent of the phage (*13*). Thus the plate stocks from the 2.8×10^7 original plaques contained $\sim 2 \times 10^7$ independent phage that expressed random sequences. Since there are 3×10^{19} different possible 15-residue peptides, any given sequence has only one chance in 2×10^{12} of occurring more than once in the library, assuming that the synthesis and resolution of the mismatched oligonucleotides is random.

We then addressed the issue of whether proteins other than antibodies to epitopes of linear peptides are able to bind to some random peptide. We screened the library for phage that bound to the biotin-binding protein, streptavidin, a protein with no known affinity for peptides. Furthermore, since one method to identify protein-binding peptides would be to screen the library with biotinylated proteins and then isolate protein-phage complexes on streptavidin-

Department of Molecular Biology, Cetus Corporation, 1400 Fifty-third Street, Emeryville, CA 94608.

*To whom correspondence should be addressed.

coated surfaces, it would be necessary to know whether phage in the library would bind directly to streptavidin.

We used streptavidin-coated polystyrene plates as an affinity matrix for examining the phage stock. Expecting a 10^3 to 10^4 enrichment during each round of selection (*4*) and since the library had a complexity of 2×10^7, we performed two rounds of selection. In the initial selection, 10^{12} phage were adsorbed to a 60-mm plate. The plates were coated with streptavidin, bovine serum albumin was then added, the phage were adsorbed for 10 min, and the plates were washed and eluted (*4*), yielding 4×10^5 phage. After preparing a plate stock (*14*) to amplify the eluted phage, we repeated the adsorption and elution from a streptavidin-coated plate, starting with 10^{10} phage and eluting 10^8 phage. We plated these phage at low density and prepared individual phage stocks from 60 randomly selected plaques (*14*).

We then determined whether the phage from these 60 isolates specifically bound to streptavidin. To avoid problems in interpreting background binding to streptavidin-coated plates, we mixed the individual phage preparations with M13mp19 phage and tested streptavidin coated plates for ability to enrich the peptide-expressing phage relative to the M13mp19 phage in the mixture. We compared the ratio of peptide-expressing phage to M13mp19 both in the initial mixture and in the eluate from the streptavidin plate.

Adsorption to and elution from streptavidin-coated plates enriched 56 of the 60 isolates by a factor of 10 relative to M13mp19. The DNA sequence analysis of 20 of the enriched isolates predicted nine different random peptide sequences, some of which occurred more than once in the 20 isolates. As examples of the nine different phage, the enrichment data for isolates A to I are shown in Table 1. For comparison, the data are shown for two isolates that failed to show enrichment and for M13LP67. If the plates were only blocked with BSA, but not coated with streptavidin, no enrichment occurred. For those phage tested, biotin (1 μM) significantly reduced the enrichment (Table 1).

Alignment of the His-Pro sequence found in the nine different predicted peptide sequences revealed an obvious consensus sequence (Table 2). The location of the consensus sequence in isolate B suggests that even the region of the random peptide closest to the proline linker is accessible for binding to foreign proteins. We have also sequenced the inserts from six random isolates that had not been selected for streptavidin binding; these predicted the expected distribution of amino acid residues and none of these predicted a His-Pro sequence. Although both histidine and biotin comprise a nitrogen and carbon containing ring, the exact nature of the interaction with streptavidin remains to be elucidated.

Table 2. The predicted sequences of the random peptides expressed by 20 streptavidin binding phage isolates. The sequences have been aligned on the common His-Pro sequence and are given in single letter code (*15*). The number of times each sequence occurred in the 20 isolates is given in the frequency column.

Isolates	Frequency	Peptide sequences
A	3	SDDWWHD HPQN LRSS
B	1	MLWYSPHSFS HPQN T
C	1	SWWLSW HPQN TKELG
D	5	ISFENTWLW HPQF SS
E	1	LC HPQF PRCNLFRKV
F	2	PC HPQY RLCQIRPLKQ
G	2	QPFL HPQG DERWYMI
H	1	ALCCLSSP HPNG AIF
I	4	LN HPMD NRLHVSTSP
Consensus		HPQn

Table 1. Enrichment of individual phage isolates on streptavidin-coated plates. For each isolate, 5 μl of phage stock (each stock had a different titer) and 5×10^9 M13mp19 phage were mixed in tris (50 mM, pH 7.5), 150 mM NaCl, bovine serum albumin at 100 μg/ml, and (for the +Biotin column) 1 μM biotin; the final volume was 50 μl. The phage from 30 μl of each mixture were then adsorbed to and eluted from streptavidin-coated 96-well plates (*4*). A reduction in the number of M13mp19 phage (blue plaques) per isolate phage (white plaques) indicated enrichment and was determined by plating both the initial mixtures and the eluates on 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (Xgal) indicator plates (*1, 6*). Phage from isolates A to I were reproducibly enriched (by a factor of 10^3 to 10^5); however, this is a qualitative assay and the significance of the differences in enrichment among the isolates A to I is not apparent. Results shown without biotin represent one of five separate experiments, results with biotin represent one of two experiments.

Isolate	M13mp19 plaques per isolate plaque					
	− Biotin			+ Biotin		
	Initial mixture	Eluate	Enrichment	Initial mixture	Eluate	Enrichment
A	28	0.002	1.4×10^4			
B	58	0.003	1.9×10^4	25	2	13
C	19	0.0007	2.6×10^4			
D	60	0.0009	6.7×10^4	320	3	107
E	140	0.012	1.2×10^4	110	1	110
F	23	0.0091	2.5×10^3			
G	28	0.0009	3.1×10^4	21	2	10
H	26	0.0024	1.1×10^4	48	1	48
I	16	0.002	8.0×10^3	17	6	2.8
Y	11	5	2.2			
Z	15	4	3.8			
M13LP67	9	5	1.8			

REFERENCES AND NOTES

1. J. Sambrook, E. F. Fritsch, T. Maniatis, Eds., *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, ed. 2, 1989).
2. G. P. Smith, *Science* 228, 1315 (1985).
3. T.-C. Lin, R. E. Webster, W. Konigsberg, *J. Biol. Chem.* 255, 10331 (1980).
4. S. F. Parmley and G. P. Smith, *Gene* 73, 305 (1988).
5. J. K. Scott and G. P. Smith, *Science* 249, 386 (1990).
6. By in vitro mutagenesis, the sequence CCGCTG starting at position 1631 in M13mp19 was converted to CGGCCG, thereby producing an Eag I restriction enzyme site without changing the amino acids encoded in this region (*1*). Similarly, the sequence TGTTCC at position 1611 was converted to GGTACC to produce a Kpn I site. A β-lactamase gene was placed in the polylinker to allow selection while propagating the vector as a plasmid; however, it also disrupted the expression of the β-galactosidase α peptide, resulting in the production of white plaques on Xgal plates (*1*). The β-lactamase gene was obtained by PCR (*16*) amplification with the plasmid pUC19 as the template with the following two oligonucleotides as primers:
(i) GCTGCCCGAGAGATCTGTATATATGAGTAAACTTGG
(ii) GCAGGCTCGGGAATTCGGGAAATGTGCGCGGAACCC.
The PCR product was digested with Bgl II and Eco RI (these sites are underlined in the primers) while double-stranded, replicative form of the phage DNA was digested with Bam HI and Eco RI. The appropriate fragments were gel-purified, ligated, and transformed into *E. coli*; cells harboring phage with the appropriate insert were selected on ampicillin plates (*1*).
7. M13LP67 DNA, which had been digested with Eag I and Kpn I, was ligated for 4 hours at room temperature and then overnight at 15°C to these two oligonucleotides (X = deoxyinosine):
(i) CTTTCTATTCTCACTCCGCTGAA(NNS)$_{15}$-CCGCCTCCACCTCCACC
(ii) GGCCGGTGGAGGTGGAGGCGG(XXX)$_{15}$-TTCAGCGGAGTGAGAATAGAAAGGTAC
The oligonucleotides were first mixed together and heated to 95°C for 5 minutes and then cooled to room temperature in 15-μl portions. The ligation mixture contained digested M13LP67 DNA (4[illegible] ng/μl), a fivefold molar excess of oligonucleotides, T4 ligase at 3.6 units/μl (New England Biolabs), 25 mM tris, pH 7.8, 10 mM $MgCl_2$, 2 mM dithiothrei-

tol, 0.2 mM ATP, and BSA at 0.1 mg/ml.
8. G. von Heijne, *Nucleic Acids Res.* 14, 4683 (1986).
9. J. F. Reidhaar-Olson and R. T. Sauer, *Science* 241, 53 (1988).
10. E. Katchalski, A. Berger, J. Kurz, in *Aspects of Protein Structure*, G. N. Ramachanaran, Ed. (Academic Press, New York, NY, 1963), p. 205.
11. W. Gray, R. S. Brown, D. A. Marvin, *J. Mol. Biol.* 146, 621 (1981).
12. 11249 cells (a *recA*, *sup*^{e}, F' derivative of MM294) were prepared for transfection by electroporation (*17*). For electroporations, 85 μl of cells (approximately 4×10^{9}) were mixed with 1 μg of M13LP67 DNA that had been resuspended in 1 mM Hepes after ligation (*7*). This mixture was then subjected to a 560-V, 5-ms pulse in a 0.56-mm gap electrode with a BTX electroporation device. Immediately after electroporation, the contents of the electrode were mixed with freshly cultured cells and plated (*1*).
13. Phage from 40 plaques were transferred by toothpick to tubes containing all the components necessary for PCR (*16*) including these two oligonucleotides:
(i) TCGAAAGCAAGCTGATAAACCG
(ii) ACAGACAGCCCTCATAGTTAGCG
After 40 cycles, the PCR products from phage with and without an insert (296 bp compared to 227 bp) could be easily distinguished on a 2 percent agarose gel.
14. N. S. Zinder and J. D. Boeke, *Gene* 19, 1 (1982).
15. Abbreviations for the amino acid residues are: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.
16. M. A. Innis *et al.*, Eds., *PCR Protocols: A Guide to Methods and Applications* (Academic Press, San Diego, CA, 1990).
17. W. J. Dower, J. F. Miller, C. W. Ragsdale, *Nucleic Acids Res.* 16, 6127 (1988).
18. We thank D. Spasic, L. Goda, and C. Levenson for oligonucleotide synthesis; J. Lee for DNA sequence analysis; D. Gelfand and F. Lawyer for providing the strain H249; G. McGregor and M. J. Polkar for literature searches; and E. McCallan for manuscript preparation.

23 March 1990; accepted 1 June 1990

Polarity and Velocity of Sliding Filaments: Control of Direction by Actin and of Speed by Myosin

JAMES R. SELLERS AND BECHARA KACHAR

Myosin filaments, which are responsible for a large repertoire of motile activities in muscle and nonmuscle cells, can translocate actin filaments both toward and away from their central bare zone. This bidirectional movement suggests that there is enough flexibility in the head portion of the tightly packed myosin molecules in the native myosin filaments to move actin filaments not only in the expected direction, but also in the direction opposite to that predicted by the regular structure of muscle—away from the center of the myosin filament.

MYOSIN FILAMENTS EXHIBIT A tight packing of the rod portion of the myosin molecules with the heads forming projections of opposite polarity on either side of a narrow central bare zone (*1*). It is commonly accepted that the cyclical interaction of myosin heads with polar actin filaments generates force that pulls the actin filaments toward the center of the bipolar myosin filament resulting in shortening of the muscle (*1*). Recently, an in vitro motility assay has been developed that allows for the visualization of the movement of fluorescently labeled actin filaments over a surface randomly coated with myosin (*2*, *3*). In this assay, the myosin molecules are not directly imaged and therefore their precise orientation is not known.

We have isolated large native thick filaments from clam adductor muscles with the use of a gentle technique (*4*). These myosin filaments can be directly visualized by video-enhanced differential interference contrast (DIC) microscopy (*5*) and appear as spindle-shaped filaments of lengths up to 20 μm (Fig. 1) (*6*). When bound to a glass surface, these myosin filaments translocated fluorescently labeled actin filaments in the in vitro motility assay (Fig. 2A). Direct comparison between the DIC and fluorescence images revealed that fluorescently labeled actin filaments could bind at any position on the myosin filament and commence directed movement (Fig. 2). Actin filaments moved both toward and away from the center of the myosin filament. Both long (2 to 8 μm) and short (<1 μm) actin filaments exhibited this behavior. The movement of actin filaments away from the center of the myosin filament is opposite that which normally occurs in muscle contraction. Analysis of the movement demonstrated that those actin filaments traveling toward the center of the myosin filament moved at a fast rate of 8.8 ± 1.4 μm/s, whereas those that were traveling away from the center moved at the much slower rate of 1.0 ± 0.3 μm/s (Fig. 2B). Some actin filaments traveled the entire length of the thick filament. In these cases the actin filament would bind to a myosin filament and commence moving at the fast rate until it crossed the bare zone, where it would slow abruptly upon encountering myosin heads of the opposite polarity. A single actin filament could reverse direction of travel by detaching and rapidly reattaching after undergoing a 180° or 360° end-to-end flip (Fig. 3A). This occurred most frequently with short actin filaments, which undergo rapid Brownian movements upon detachment. The direction of travel after the flip was correspondingly recovered or reversed, consistent with the intrinsic polarity of the actin filament. On some occasions a longer actin filament sliding off the end of the myosin filament reattached through its initial leading end and proceeded to move back down the same myosin filament in the opposite direction. Two actin filaments traveling in opposite directions on a single myosin filament could pass one another unimpeded, which is not surprising given the relatively large diameters of the myosin filaments (Fig. 3B). When this occurred one filament would be traveling at the fast speed and the other at the slower speed.

We also observed that the actin filament was very flexible and behaved more as a rope than as a rigid rod when moving in this assay. When a long actin filament traveling at the fast rate crossed into the region of opposite polarity of the myosin filament, its leading end slowed abruptly while the tail end, which was still in contact with myosin heads of the correct polarity, continued to move fast. Because of the difference in the sliding speeds of each end of the actin filament, it formed a flexible loop in the middle portion that was over the bare zone and was free of attachment (Fig. 3C). Actin filaments could also interact simultaneously with more than one adjacent myosin filament, producing independent sliding actions and often undergoing large changes in angles (Fig. 3D). The implications of this flexibility may be more commonly manifest in nonmuscle systems where the actin filaments are typically much longer than the myosin filaments and where a single actin filament may possibly interact with more than one myosin filament (*7*). This flexible nature of the long actin filaments indicates that the actions of well-separated myosin heads are not necessarily integrated along the actin polymer chain. Actin filaments often exist in bundles within cells. Such bundles may allow for a structure with more

J. R. Sellers, Laboratory of Molecular Cardiology, National Heart, Lung and Blood Institute, National Institutes of Health, Bethesda, MD 20892.
B. Kachar, Laboratory of Molecular Otology, National Institute on Deafness and Other Communication Disorders, National Institutes of Health, Bethesda, MD 20892.